PERFORMANCE EVALUATION OF MULTIRESOLUTION TEXTURE ANALYSIS OF STEM CELL CHROMATIN

Rami Mangoubi[1]*, Mukund Desai*[1]*, Nathan Lowry*[1]*, and Paul Sammak*[2]

[1]C.S. Draper Laboratory, Cambridge, Massachusetts, 02139
[2]Magee-Womens Research Institute,
University of Pittsburgh, Pittsburgh, PA 15213

ABSTRACT

We apply texture image analysis to automated classification of stem cell nuclei, based on the observation that chromatin in human embryonic stem cells becomes more granular during differentiation. Using known probability models for texture multiresolution decompositions, we derive likelihood ratio test statistics. We also derive the probability density functions of these non-Gaussian statistics and use them to evaluate the performance of the classification test. Results indicate that the test can distinguish with probability 0.95 between nuclei that are pluripotent and those with varying degrees of differentiation. The test recognizes nuclei with similar differentiation level even if prior information says the contrary. This approach should be useful for classifying genome-wide epigenetic changes and chromatin remodeling during human development. Finally, the test statistics and their density functions are applicable to a general texture classification problem.
Index Terms: texture, classification, stem cell, non Gaussian

1. INTRODUCTION

Biologists and microscopists routinely characterize chromatin condensation by biochemical methods that require destruction of the specimen, which prevents evaluation of chromatin dynamics and limits the use of complementary tests on the same sample. Here, we evaluate a method of measuring chromatin compaction and organization using the live cell reporter, GFP H2B. We evaluate the differences in chromatin condensation as pluripotent human embryonic stem cells undergo neuronal differentiation, and in addition evaluate the dynamic changes in single nuclei over time. In [5, 7, 10] a state of the art methodology is proposed for characterizing the characteristics of stem cell colonies and nuclei using texture based image analysis. The intent of the methodology, which has been validated on several specimens, is to provide a decision aid to the microscopist. The expectation is that a sound image texture analysis methodology will be rapid, accurate, and consistent, without altering the specimen under examination.

Work supported by NIH grant 1 R01 EB006161-01A2 and C.S. Draper Laboratory, Cambridge Massachusetts.

The methodology used in [5, 7, 10] is based on a texture analysis approach suggested in [3] that combines multiresolution analysis and statistical modeling. Specifically, it is assumed that a wavelet decomposition of a texture class of interest yields random coefficients that possess a generalized Gaussian probability density function (pdf). Hypothesis testing detection problems based on the generalized Gaussian densities have also been considered in [1], and applied to other biomedical applications, specifically functional MRI [2].

In [3], an approximate textural library retrieval scheme that uses the Kullback-Leibler (KL) distances is developed. As discussed in [3], the KL distance is equivalent to the maximum likelihood approach when the sample size is small or even large but finite only under certain restrictive conditions. Moreover, correct classification performance is difficult to obtain using the KL distance.

In this paper, we build upon and extend the work in the above mentioned references by deriving likelihood ratio statistics and their respective distributions for texture analysis. The distributions can be used to obtain the performance of the likelihood ratio based classifier. We show the derivation of the statistics and their density functions in the next section. We have validated this approach on several applications, including generic textures, egg cells, stem cell colonies and individual nuclei. We describe one such application in Section 3, and we illustrate the performance of the classifier on this application in Section 4. Section 5 presents conclusions.

2. FORMULATION AND APPROACH

We first summarize the statistical nature of the decomposition as established in [3]. Next, we formulate a likelihood ratio test, derive the statistics, and their density functions.

2.1. Statistical Properties of Texture Wavelet Coefficients

It is shown in [3] that the values of the coefficients at each wavelet subband have a *generalized Gaussian* distribution, or

$$f(x;\omega,p) = \frac{p}{2\omega\Gamma(1/p)} e^{-(|x|/\omega)^p} \qquad (1)$$

where x is the random variable for a particular wavelet subband of the texture, ω is a width parameter proportional to the standard deviation, and p is a tail thickness shape parameter. A third parameter, the location or mean, is found to be zero in (1) for the subbands of interest. The term $\Gamma(1/p)$ represents the Gamma function. Note that when $(p, \omega) = (2, \sqrt{2})$, we have a standard Gaussian density. Each texture would have a representative density function (1) at each decomposition level. Thus, given a texture decomposition that provides B wavelet bands, the characterizing pdf for that texture is

$$f(x_1, \ldots, x_B; \omega_1, \ldots, \omega_B, p_1, \ldots, p_B) = \prod_{b=1}^{B} \frac{p_b}{2\omega_b \Gamma(1/p_b)} e^{-(|x_b|/\omega_b)^{p_b}} \quad (2)$$

The above texture characterization is an approximation; it assumes that the wavelet coefficients are statistically independent across resolution levels. Though it may be approximate, this statistical model of texture decomposition has been validated on many types of textures [3]. In particular, it was shown to be applicable to the analysis of stem cell colonies in [5, 7] and to nuclei of various stem cell lineages in [10].

In all these references, the Kullback-Leibler distance is used as a basis for classifying the texture. The KL distance between two density functions f_1 and f_2 is given by

$$D_{KL}(1, 2) = \int f_1(x) \ln \left(\frac{f_2(x)}{f_1(x)} \right) dx. \quad (3)$$

That is, given a new texture that needs to be assigned to one class from among C classes, we compute the wavelet coefficients from samples for the new texture, and we use these sample decompositions to estimate the generalized Gaussian parameters [3]. We now have knowledge of the new texture's pdf, f_{new}, and can select the class c^* whose pdf has the shortest KL distance from the new texture's pdf

$$c^* = \arg\min_c D_{KL}(f_{\text{new}}, f_c) \quad (4)$$

As noted in [3], KL distance based classification is equivalent to maximum likelihood classification only in the asymptotic sense. For finite samples, however, this equivalence does not hold, except for the unrealistic case where the shape parameters have the same value in (2), when $p_1 = p_2 = \ldots = p_B$.

2.2. The Likelihood Ratio Test

We now formulate the likelihood ratio test and derive its statistic for the purpose of classifying textures. The pdf of this statistic is derived next and is used in Section 4 for evaluating the performance of maximum likelihood texture classification as applied to the evaluation of stem cell pluripotency level.

Consider the problem of assigning a texture to one of two classes characterized by different generalized Gaussian density parameters for their wavelet decomposition. Assume we have B subbands and S samples for this texture. Denote the associated random variables by $x_{sb}, s = 1, \ldots, S$, and $b = 1, \ldots, B$. As in [3], we assume that the subband decompositions are independently distributed. Let X_s be the vector of random variables from each of these samples $s = 1, \ldots, S$. We formulate the following binary hypothesis test

$$H_0 : X_1, \ldots, X_b, \ldots X_S \sim f_0 = \prod_{b=1}^{B} \prod_{s=1}^{S} f_{0b}(x_{sb}) \quad (5)$$

$$H_1 : X_1, \ldots, X_b, \ldots, X_S \sim f_1 = \prod_{b=1}^{B} \prod_{s=1}^{S} f_{1b}(x_{sb}) \quad (6)$$

where each of the densities p_{0b} and p_{1b}, $b = 1, \ldots, B$ is a generalized Gaussian density function given in (1), with respective parameters $(\omega_1, p_1), \ldots, (\omega_B, p_B)$.

2.2.1. The χ^p random variables

To derive the log-likelihood ratio statistic in Section 2.2.2 for the above test and its density functions, we need to generalize the χ^2 random variable. For a generalized Gaussian random variable x with parameters ω and p, define

$$\chi^p = \left| \frac{x}{\omega} \right|^p \quad (7)$$

For $p = 2$, we have the χ^2 random variable, which is the square of a standard normal variable. Likewise, χ^p is a generalization of χ^2 with respect to the generalized Gaussian variable with width parameter $\omega = 1$, raised to the power p, whose pdf is

$$f_{\chi^p}(x) = \frac{1}{\Gamma(1/p)} e^{-x} |x|^{-1+1/p}, \quad x \geq 0 \quad (8)$$

The χ^2_N, or the χ^2 random variable with N degrees of freedom, is simply the sum of N independent χ^2 random variables. Likewise, we can define the χ^p_N random variable to be the sum of N independent χ^p random variables.

To go a step further, if the shape parameters are different for each of the random variables in the sum, we can define, with $\overrightarrow{p} = (p_1, \ldots, p_N)$, the random variable

$$\chi^{\overrightarrow{p}} = \sum_{i=1}^{N} \chi_i^{p_i} \quad (9)$$

The density function of the above random variable is obtained by convolving the density functions of (8)

$$f_{\chi^{\overrightarrow{p}}}(x) = \frac{1}{\Gamma(\sum_{i=1}^{N} 1/p_i)} e^{-x} |x|^{-1+\sum_{i=1}^{N} 1/p_i}, \quad x \geq 0 \quad (10)$$

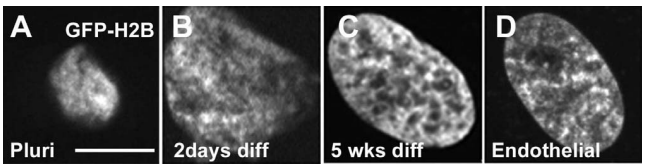

Fig. 1. (A-C) with increasing chromatin granularity, compared to a control somatic endothelial cell (D). Bar in A is 10 μm.

2.2.2. The likelihood ratio statistic

The log-likelihood ratio for the hypothesis test (5-6) is expressed in terms of these new random variables. With $N = SB$, where S is the number of samples for the texture and B the number of wavelet bands per sample, we have

$$\begin{aligned} \Lambda(X_1, \ldots, X_s) &= \ln \frac{\prod_{b=1}^{B} \prod_{s=1}^{S} f_{1b}(x_{sb})}{\prod_{b=1}^{B} \prod_{s=1}^{S} f_{0b}(x_{sb})} \\ &= \sum_{i=1}^{N} \chi_{i0}^{p_{0i}} - \chi_{i0}^{p_{1i}} + K \\ &= \chi_0^{\vec{p}_0} - \chi_1^{\vec{p}_1} + K \end{aligned} \tag{11}$$

where K is a constant dependent on the scale and shape parameters. Now we have the test

$$\chi_0^{\vec{p}_0} - \chi_1^{\vec{p}_1} \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{12}$$

Here, T denotes the classification threshold that absorbs K. The above methodology has been validated on generic textures as well as images of various stem cell nuclei. In the next section, we describe one such application.

3. APPLICATION TO STEM CELL NUCLEI

Human embryonic stem cells (hESC, line UC06 from the NIH-approved registry) were grown under standard conditions on mouse feeder cells. Pluripotency of hESC was routinely confirmed by immunostaining for the pluripotency marker, Oct-4. hESCs were induced to differentiate for up to 5 weeks by plating on feeder cells at half the normal density, which induced differentiation to early neuronal lineages as determined by the neural marker, nestin [8]. We visualized chromatin in living cells with a fluorescent histone that bound to DNA. Cells were transiently transfected with a plasmid expressing the histone H2B labeled with the fluorescent protein GFP [6]. 4-D movies were acquired with a spinning disk microscope (Perkin Elmer) using a 40x 1.3NA Nikon objective with a resolution of 0.2 μm. We observed that nuclei in pluripotent cells were small and chromatin was generally smooth textured (Figure 1 A). During differentiation (Figure 1 B) we found

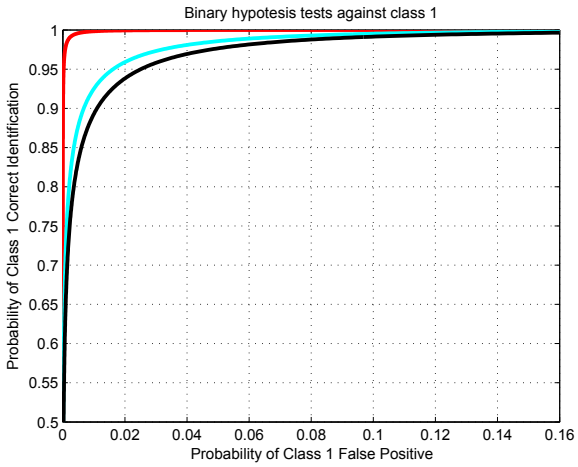

Fig. 2. Probability of correctly identifying a pluripotent nucleus (class 1) vs. probability of a false positive from partially or totally differentiated nuclei.

that chromatin became more granular and did not vary over time, unlike pluripotent cells. By 5 weeks (Figure 1 C), differentiated stem cells were nearly as granular as an adult human vascular endothelial cell (Figure 1 D). Pluripotent nuclei are physically very plastic and become less compliant during differentiation due in part to chromatin condensation [9]. The bright fluorescent regions within the nuclei that we observed reflect compact chromatin supercoiling which limits accessibility of DNA to soluble proteins [4]. Chromatin condensation is biologically significant because transcription factors and activators need to have access to DNA in order to express genes. The granularity of chromatin therefore reflects the segregation of the nucleus into domains of high density (bright areas, heterochromatin) and low density (dim areas, euchromatin). Since heterochromatin generally contains silenced genes, texture analysis provides a direct measure of the degree of gene silencing by chromatin remodeling.

4. PERFORMANCE RESULTS

Figure 1 shows nuclear images of 4 cells at first time in a time-lapse series of 9 images over a 10 minutes duration. The last two classes are very close and are expected to be indistinguishable. Of utmost importance is the ability to identify totally pluripotent nuclei, while minimizing false positives. Figure 2 plots the probability of correctly identifying such a nucleus, against the probability of misclassifying a differentiated nucleus from each of the other three classes as pluripotent. Each differentiated class is shown by a separate curve. It is shown, for instance, that for a probability of misclassification of less than 0.05, we have a correct classification of larger than 0.95, for any of the three alternate classes.

Pairwise comparison of cells from various classes is shown

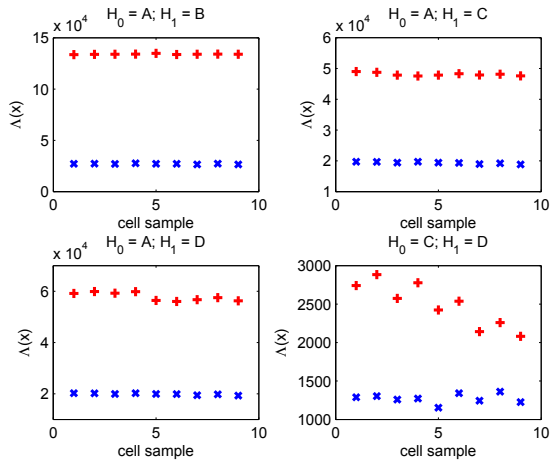

Fig. 3. Pairwise hypothesis testing between classes A and B (top left), A and C (top right), A and D (bottom left), and C and D (bottom right). Cells for which the null hypothesis H_0 is true are plotted in blue, while cells for which the alternative hypothesis H_1 is true are plotted in red. Results indicate that the likelihood ratio statistics show clear separation between pairs of classes.

in Figure 3. The marker GFP-H2B is imaged in each of the 4 cells shown in Figure 1 at 9 time points (Pluripotent cell A; differentiated cell B; differentiated cell C; control endothelial cell D). Because the cells from each class carry the same color, and cells are segregated by colors, the results show that the likelihood ratio test's statistic described in (12) enables clear separation between cells of distinct classes.

5. CONCLUSION

Results indicate that the likelihood ratio test and its derived density function is a promising approach for texture classification since it enables the evaluation of classification performance. We have applied this test to several types of texture, and as the results show for a class of stem cell nuclei, the test can be a decision aid to the microscopist when evaluating cell characteristics, for instance stem cell pluripotency. Reorganization of the genome occurs during embryonic development, tumorigenesis, viral infection, DNA repair and apoptosis. The ability to measure statistically significant changes in nuclear texture will provide new ways of measuring these phenomena. Current research focuses on applying the above test to other cell cultures and on establishing high speed, reliable, and widely applicable classification methods.

6. REFERENCES

[1] M. Desai, R. Mangoubi, "Robust Gaussian and non-Gaussian matched subspace detection", *IEEE Trans. on Sig. Proc.*, Pp.3125-27.

[2] M. Desai, R. Mangoubi, J. Shah, W. Karl, D. Kennedy, H. Pien, and A. Worth, "Functional MRI activity characterization using curve evolution," *IEEE Trans. on Medical Imaging* , Vol.21, No. 11, Nov, 2002, pp. 1402-1412.

[3] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, Vol. 11, no. 2, , Feb. 2002, pp. 146–158.

[4] S. M. Gorisch, M. Wachsmuth, et al. "Histone acetylation increases chromatin accessibility", *J. Cell Science*, Vol. 118, Pt. 34, 2005, pp. 5835-34.

[5] C. Jeffreys, "Support vector machine and parametric wavelet-based texture classification of stem cell images," M.S., Massachusetts Institute of Technology, Cambridge, MA, June 2004.

[6] T. Kanda, K. Sullivan, et al. "Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells" , *Curr. Biology* Vol. 8, No. 7, 1998, pp. 377-85.

[7] R. Mangoubi, C. Jeffreys, A. Copeland, M. Desai, and P. Sammak, "Non-invasive image based support vector machine classification of human embryonic stem cells," *Proc. IEEE Int'l Symp. on Biomed. Imaging*, Washington, D.C., April, 2007.

[8] J. A. Ozolek, E. P. Jane, et al. "Human embryonic stem cells (HSF-6) show greater proliferation and apoptoses when grown on glioblastoma cells than mouse embryonic fibroblasts at day 19 in culture: comparison of proliferation, survival, and neural differentiation on two different feeder cell types." *Stem Cell Dev.* Vol. 16, No.3, 2007, pp.403-12.

[9] J. D. Pajerowski, K. N. Dahl, et al. "From the Cover: Physical plasticity of the nucleus in stem cell differentiation." *Proc. Natl. Acad. Sci. USA* , Vol 104, No.40, 2007, pp.15619-24.

[10] P. Sammak, V. Abraham, R. Ghosh, J. Haskins, E. Jane, P. Petrosko, T. Erb, T. Kinney, C. Jeffreys, M. Desai, and R. Mangoubi, "High Content Analysis of Human Embyonic Stem Cell Growth and Differentiation," Ch. 9. in *High Content Screening*, S. Haney, ed., John Wiley and Sons, inc. To appear in 2008.

[11] N. R. C. Xu, S. Police and M. K. Carpenter, "Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells," *Circ Res*, vol. 91, no. 6, pp. 501–8, 2003.